

Welcome to the Data Engineer Career Path

Data Engineers vs Data Scientists

Data scientists and data engineers often work closely together. Data scientists work on the analytics side of this partnership, creating data models and analytics studies. Once a data model or study is prepared by data scientists, data engineers **deploy** it by creating automated data pipelines that

- regularly update the data
- test for and log any errors
- load the output to a cloud database or business intelligence tool

Data Engineers as Librarians

Data Engineers are often compared to librarians: they won't do your research for you, but they will make sure the resources you need are properly cataloged and accessible. When a team at a company needs data, it is the data engineer's job to make sure the data they need exists and is organized in a database or business intelligence tool.

Data Engineering Tools

Data Engineers use programming languages like Python and SQL to work with data. To automate processes, they also often work with the command line interface, a tool for sending commands directly to a computer.

Python

Python is a general-purpose computer programming language that has become one of the most common programming languages in the data world. Most data scientists and engineers who use Python work with pandas, a set of special commands developed in Python that make handling data easier and more efficient.

SQL

SQL (Structured Query Language) is a programming language designed specifically for working with databases. Programs written in SQL are called queries since they are often used to ask for information from databases. But SQL can also be used to create new tables of data or restructure existing tables.

Cloud Deployment

Cloud deployment is the process by which data engineers move data onto specialized database servers accessible over the internet.

 **Print**  **Share** ▼